Parsing with Analogical Substitutions: An Exemplar-based Model of the Emergence of Syntactic Structure

MOTS-CLÉS : inférence analogique, émergence, théorie des exemplaires, apprentissage distributionnel.

KEYWORDS: analogical inference, emergence, exemplar theory, distributional learning.

1 Introduction

Exemplar-based inference methods in linguistics make predictions about novel linguistic expressions by comparing them to attested expressions, or *exemplars*, such that a novel expression is predicted to behave like its most similar exemplars. Exemplar-based linguistic theories represent linguistic knowledge as a database of exemplars plus some exemplar-based inference method — notably, generalizations over exemplars are not explicitly represented but are implicit in the computation of a prediction, as opposed to rule-based theories (where generalizations are represented as rules) and connectionist theories (where generalizations are represented as the weighted connections of a neural network).

While recent advances in large language models have demonstrated the applicability of connectionist theories in natural language processing tasks, exemplar-based theories remain theoretically attractive for their interpretability. Whenever an exemplar-based inference method makes a prediction, it inherently identifies a similarity-weighted set of exemplars reflecting the contribution of each exemplar to the predicted behavior, allowing us to form hypotheses about why expressions behave the way they do.

In any exemplar-based theory of *parsing* (assigning syntactic structures to sentences), the exemplars that contribute to the predicted structure would intuitively include sentences that are grammatically similar to the input sentence. A recent example is Ambridge's (2019, p. 534) account of exemplar-based sentence processing, where a passive sentence like *the vase was broken by the hammer* is said to evoke, among other things, passive exemplar sentences like *the window was smashed by a ball*.

But neither Ambridge's account nor any other exemplar-based account of sentence processing (that we are aware of) includes an algorithm that parses an input sentence by finding the most grammatically similar exemplar sentences. In our presentation we will propose such a method.

2 Recursive analogical substitutions

Our parsing method is the byproduct of a solution to the problem of finding, given a possibly unattested phrase (string of words), the corpus phrases that can best predict its occurrence in arbitrary contexts, which we call its *analogical phrases*. We state the problem, describe our solution to it and then formulate our solution as a parsing method. We make a background assumption: we never come across words that did not occur in our training corpus.

Let s be a possibly unattested phrase. Our task is to find corpus phrases s' that are *substitutable* by s to a high degree with respect to grammaticality, where the degree of substitutability of s' by s is defined as the degree of certainty with which we can assume that s can grammatically occur in a context given that s' can grammatically occur in that context.

We give our solution as a recursive algorithm. The base case is that s occurs enough times in our corpus so that we can find the substitutable phrases s' simply by comparing their distributions to s. We will describe our preliminary solution to the base case in the presentation (and mention some of the solutions already present in the literature, e.g. Redington et al. (1998) and Walsh et al. (2010)). For now, let us assume that we have solved the base case by using some method to return corpus phrases s' with a high degree of substitutability by s. For the sake of conciseness, if s' is substitutable by s to a high degree, throughout this paper we will write $s' \Rightarrow s$ (expressing that we can "rewrite" s' as s) and simply say that s' is substitutable by s. (In our presentation we will define the degree of substitutability as a continuous variable.)

The recursive case is that s is unattested or there is too little distributional information about it for our solution above to find phrases other than s itself that are substitutable by it. If s is a single word, we just return it — there is no way for us to find other substitutable corpus phrases. Otherwise s is a multi-word phrase: let us proceed to describe this case with $s =$ *my friend saw the rainbow* and illustrate the process in Figure 1 below.

We split s up into two possible constituents $s_1 =$ *my friend* and $s_2 =$ *saw the rainbow*, and recursively call our algorithm on them to get corpus phrases $a_1 =$ *our teacher* and $a_2 =$ *heard it* such that $a_1 \Rightarrow s_1$ and $a_2 \Rightarrow s_2$. We then apply our base case algorithm to a_1 and a_2 to find corpus phrases $s'_1 =$ *she* and $s'_2 =$ *arrived* such that $s'_1 \Rightarrow a_1$, $s'_2 \Rightarrow a_2$ and $s'_1 s'_2$ occurs in the corpus. For each possible split of s we collect these concatenations $s'_1 s'_2$ and return them as its analogical corpus phrases.

The assumptions that conceptually underlie the recursive process are that substitutability is *transitive* and *compatible* with the concatenation operation, the latter meaning that for any strings s_1, s_2, s'_1, s'_2:

$$\text{if } s'_1 \Rightarrow s_1 \text{ and } s'_2 \Rightarrow s_2, \text{ then } s'_1 s'_2 \Rightarrow s_1 s_2.$$ [1]

[1] If we understood "⇒" as *full* substitutability, both properties could be proved instead of having to be assumed; but we understand "⇒" as a *high degree* of substitutability, in which case neither can be proved.

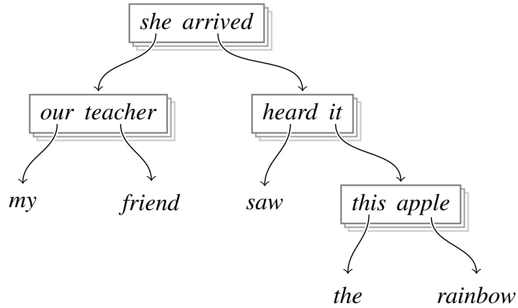

Figure 1: Finding analogical evidence for the sentence *my friend saw the rainbow* through recursive analogical substitutions. (Arrows indicate substitutability, e.g. *she* ⇒ *our teacher*; stacks indicate multiple corpus phrases.)

E.g. in Figure 1 we deduce *she* ⇒ *my friend* in two steps: (1) we use compatibility to deduce *our teacher* ⇒ *my friend* from *our* ⇒ *my* and *teacher* ⇒ *friend*, and (2) we use transitivity to deduce *she* ⇒ *my friend* from *she* ⇒ *our teacher* and *our teacher* ⇒ *my friend*.

Figure 1 illustrates the inference process for a particular choice of bracketing for the sentence *my friend saw the rainbow*; we might find different analogical phrases if we split up e.g. the phrase *saw the rainbow* into the phrases *saw the* and *rainbow*. In fact, we predict that we would find *fewer* analogical phrases using this latter bracketing, since there is no word that is substitutable by the phrase *saw the* to the degree that e.g. the word *it* is substitutable by the phrase *the rainbow*.

In general, we can evaluate a binary bracketing of a phrase by the degrees of substitutability of the analogical phrases we can find through it, and take the best bracketing as the parse tree of the phrase — in essence, we can think of the syntactic structure of a phrase as a trace of the recursive analogy-finding process. Our theory of language thus gives rise to syntactic structure naturally, without resorting to stipulation: our thesis is that the syntactic structure of a phrase emerges from the process of analogical inference from exemplars.[2]

Implementations of our analogical phrase finding algorithm have only occasionally yielded intuitively acceptable parse trees; we consider this algorithm a starting point for developing an explicitly defined exemplar-based theory of language. In our presentation we will describe our current implementation of the parsing method and present example parses.

[2]The best parse tree of a phrase may coincide with the standard syntactic structure associated with the phrase, as in Figure 1, or it may not.

References

AMBDRIDGE, B. (2019). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, **40**, 509–559. DOI: https://doi.org/10.1177/0142723719869731

REDINGTON, M., CHATER, N., & FINCH, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**, 425–469. DOI: https://doi.org/10.1207/s15516709cog2204_2

WALSH, M., MÖBIUS, B., WADE, T., & SCHÜTZE, H. (2010). Multilevel exemplar theory. *Cognitive Science*, **34**, 537–582. DOI: 10.1111/j.1551-6709.2010.01099.x